REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b RESTRICTIVE MARKINGS
Unclassified	

2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION
Institute for Brain and Neural Systems		Personnel and Training Research Programs Office of Naval Research (Code 1142PT)

6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
Brown University Providence, Rhode Island 02912	800 North Quincy Street Arlington, VA 22217-5000

8a. NAME OF FUNDING / SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
		N00014-91-J-1316

8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS			
Brown University Providence, RI 02912	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)

Classification of Underwater Mammals using Feature Extraction Based on Time-Frequency Analysis and BCM Theory

12. PERSONAL AUTHOR(S)

Quyen Q. Huynh, Leon N Cooper, Nathan Intrator and Harel Shouval

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical Report	FROM _____ TO _____	May, 1996	13

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
05	08		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Underwater mammal sounds classification is demonstrated using a novel application of wavelet time/frequency decomposition and feature extraction using the BCM neuron. The system achieves outstanding classification performance even when tested with mammal sounds recorded at very different locations (from training).

19961105 040

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dr. Joel Davis	(703) 696-4744	

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted.
All other editions are obsolete.

DTIC QUALITY INSPECTED 3

Classification of Underwater Mammals using Feature Extraction Based on Time-Frequency Analysis and BCM Theory

Quyen Q. Huynh*   Leon N Cooper   Nathan Intrator[t]   Harel Shouval

Physics Department and
Institute for Brain and Neural Systems
Brown University

May, 1996

Abstract

Underwater mammal sounds classification is demonstrated using a novel application of wavelet time/frequency decomposition and feature extraction using the BCM neuron. The system achieves outstanding classification performance even when tested with mammal sounds recorded at very different locations (from training).

1 Introduction

Detection, classification and localization (DCL) are among the most important and challenging goals of underwater signal analysis. A cocktail of sounds which includes biological sounds (dolphins, sperm whales, shrimp etc.) is mixed with environmental sounds (estuaries, crackings of ice, rain) and man made sounds (torpedoes, submarines, surface ships) dramatically reduces recognition performance.

It is well known that the features presented to a classifier play a crucial role on its performance. Indeed, the feature set selected may be more important than the classifier architecture itself. Recently, with the tremendous advances in time-frequency analysis (wavelet packet, local trigonometric basis, Gabor expansions), different feature extraction methodologies [5, 18, 12] have been proposed, based on the localization properties of the time-frequency basis functions. It has been shown that using a wavelet representation of the acoustic signals, one can achieve improved classification [12]. This has led to the increased interest in methods for feature extraction from this data representation.

Wavelet representation is merely a different full representation of the same signal. While it suggests natural ways to reduce representation dimensionality by keeping only the highest energy coefficients (similar to keeping only the first few Principal Components or Fourier coefficients of the signal), there is no rigorous result showing that these will be a useful representation for the purpose of signal classification and detection. The need for dimensionality reduction is clear; It

---

*On leave, Naval Undersea Warfare Center, Newport, RI 02841.
[t]On leave, Computer Science Department, Tel-Aviv University

follows from the *curse of dimensionality* [2] namely, the fact that the number of data points needed for a robust parameter estimation of the data density grows exponentially with the dimensionality. The problem of feature extraction is fundamental in information science. One looks for an efficient and compact representation of data which leads to new insight into the problem to be solved. Under some conditions, features extracted with an unsupervised learning procedure may be more robust and general than those extracted by a supervised learning procedure. This is because the unsupervised algorithm must focus on the underlying structure of the data and not on pre-assigned labels which may not reveal the full structure of the data (especially under small training set). The BCM theory was developed to understand and model the plasticity of the mammalian visual cortex. This model has recently been extended to a lateral inhibition network [14] and a statistically motivation variant of it has been used in various high dimensional feature extraction tasks [15, 17].

In this paper, we use a network of BCM neurons for optimal feature extraction from a wavelet representation, leading to improved classification of underwater acoustic signals. We emphasize here that the BCM network is not playing the role of a classifier; rather, its role is feature extraction.

1.1 Feature Extraction from Wavelet Representations

Previous approaches to feature extraction from wavelet representation were based on signal energy [5, 18, 12]. While this is not necessarily the best statistic of the signal for the purpose of classification, it was a must in the methods that have been used for feature extraction; In [5] and [18], the training set was analyzed using the time-frequency energy map of the wavelet packet decomposition tree. Coifman and Saito [5] used statistical considerations to determine the optimal wavelet packet basis for classification, which they termed the "local discriminating basis" (LDB). Unknown signals were then projected onto this LDB and classification of the unknown signals was based on the time-frequency coefficients of only those basis functions in the LDB with the largest "discriminating power." Willsky et al. [18] determined relevant features from a **time-averaged** energy map, not necessarily corresponding to a single wavelet packet basis. For each signal class in the training set, an energy matrix was constructed and the singular vectors of this matrix were used to identify the dominant energy pattern of each class. The features were then selected from the energy bins of the wavelet packet basis which corresponded to the peak values of the "primary singular" vectors. Huynh et al. [12] approached the binary classification problem by searching the wavelet packet library for another "discriminating basis" (LDB-2), using the "best basis" paradigm of Coifman and Wickerhauser [6] to find the basis that best approximated the **difference** of the two classes of signals. LDB-2 was thus the basis which maximized the **separation** of the two classes. Unknown signals were then projected onto the LDB-2 and classified by feeding a fixed number of the largest time-frequency coefficients of the LDB-2 (along with their corresponding time and frequency indices) into a standard classifier such as the back propagation artificial neural network (ANN) [19].

2 Projection Index for Classification: The Unsupervised BCM Neuron

Exploratory projection pursuit theory [11, 10] tells us that search for structure in input space can be approached by a search for deviation from normal distribution of the projected space[1]. Furthermore, when input space is clustered, a search for deviation from normality can take the form of search for multi-modality, since when clustered data is projected in a direction that separates at least two clusters, it generates multi-modal projected distributions.

It has been recently shown that a variant of the Bienenstock, Cooper and Munro neuron (BCM) [1] performs exploratory projection pursuit using a projection index that measures multi-modality [14]. This neuron allows modeling and theoretical analysis of various visual deprivation experiments [14] and is in agreement with the vast experimental results on visual cortical plasticity [4]. A network implementation which can find several projections in parallel while retaining its computational efficiency, was found to be applicable for extracting features from very high dimensional vector spaces [16, 13].

In the single neuron case, the neuronal activity (in the linear region) is given by $c = m \cdot d$. where d is the input vector and m is the synaptic weight vector (including a bias). The essential properties of the BCM neuron are determined by a modification threshold Θ_m (which is a nonlinear function of the history of activity of the neuron) and a ϕ function that determines the sign and amount of modification Θ_m. The synaptic modification equations are given by

$$\frac{dm_i}{dt} = \mu \; \phi(c, \Theta_m)d_i,$$

where in a simple form $\Theta_m = E[(m \cdot d)^2]$ and $\phi(c, \Theta_m) = c(c - \Theta_m)$.

In the lateral inhibition network of nonlinear neurons the activity of neuron k is given by $c_k = m_k \cdot d$, where m_k is the synaptic weight vector of neuron k. The *inhibited* activity and threshold of the $k'th$ neuron is given by

$$\tilde{c}_k = \sigma(c_k - \eta \sum_{j \neq k} c_j),$$
$$\tilde{\Theta}_m^k = E[\tilde{c}_k^2],$$

for a monotone saturating function σ.

The projection index for a single neuron is given by

$$R(w_k) = -\{\frac{1}{3}E[\tilde{c}_k^3] - \frac{1}{4}E^2[\tilde{c}_k^2]\}.$$

The total index is the sum over all neurons in the network. The resulting stochastic modification equations for a synaptic vector m_k (the negative gradient of the index) in the network are given by:

$$\dot{m}_k = \mu[\phi(\tilde{c}_k, \tilde{\Theta}_m^k)\sigma'(\tilde{c}_k) - \eta \sum_{j \neq k} \phi(\tilde{c}_j, \tilde{\Theta}_m^j)\sigma'(\tilde{c}_j)]d.$$

---

[1]In a neural net architecture this is the space generated by the hidden unit activity of the feed forward network.

3

This network is a first order approximation to a lateral inhibition network (using a single step relaxation). Its properties and connection to a lateral inhibition network as well as some related statistical and computational issues are discussed in [14].

Under reasonable assumptions, the BCM algorithm (with k BCM neurons) produces k weight vectors which converge iteratively to fixed points corresponding to states of "maximum selectivity." In other words, for a single BCM neuron, the converged weight vector becomes orthogonal to all cluster centers except one. The feature set of the BCM algorithm is formed by the convolutions of the k weight vectors with the unknown data.

Lateral inhibition in the network allows the construction of an array of feature-selective cells in which the same feature is not selected more than once and all features of the data set are represented in an orderly fashion.

3 Feature Extraction Based on Time-Frequency Analysis and BCM Theory

Our previous works [12] on using wavelet transforms for feature extraction have shown good results in the classification of marine mammals (dolphins, sperm whales and porpoises). Modern time-frequency techniques (wavelet packet, local trigonometric basis, Gabor expansions) are considered as tools for providing an efficient data representation to transform the original data set to a preliminary feature set. However, the *curse of dimensionality* [2] suggests that classification may be improved if a dimensionality reduction takes place before the classification stage. In this case, applying the BCM algorithm to the preliminary feature set (time-frequency-transformed data) reveals the important clues of the underlying structure of the data. The use of wavelet representation is supported by the fact that classification results obtained by feature extraction from the raw signal are worse than those obtained from the wavelet representation (Table 2).

We approach the problem of building a global and robust classifier that combines the virtues of modern adaptive time-frequency techniques and BCM optimal selectivity as follows:

1. Choose an efficient coordinate system (library of orthogonal and nonorthogonal bases) to transform the original data set to a preliminary feature space.

2. Construct a network of connected k BCM neurons with lateral inhibition.

3. Train the k BCM neurons on the transformed data to produce k stable weight vectors.

4. Extract k crucial features which are the convolution outputs of the k weight vectors with the transformed unknown data.

5. Present the k features as inputs to a classifier e.g. the back propagation classifier [19].

3.1 Signal description

The types of signals explored in this study are the marine mammal sounds namely porpoise and sperm whale which were recorded at a sampling rate of 25 kHz at various locations such as the Gulf of Maine, the Mediterranean and the Caribbean sea. We consider large original data files where sounds consist intermittently of mammal sounds and background noise. Note that each of these
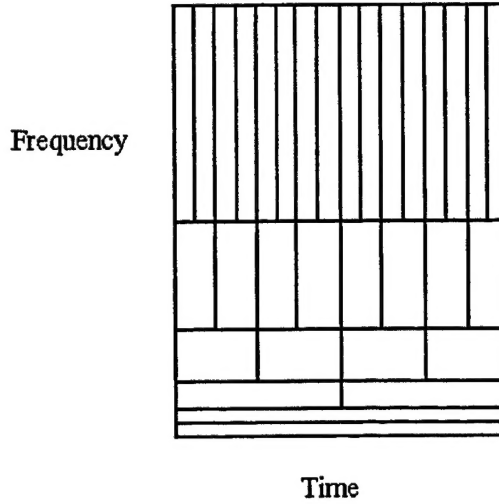
Time-Frequency wavelet representation

Figure 1: Dyadic time-frequency tiling of the phase plane. The frequency axis is partitioned in an octave-band fashion. Low frequency band with low temporal resolution is at the bottom, while higher frequency bands with high temporal resolution are towards the top of the figure. The entire phase plane is covered by disjoint rectangles of equal area. On the time-frequency plane, the highest frequency bin was 6.25 - 12.5 kHz and there were 16384 wavelet coefficients spanning over the bandwidth of the signals in the time domain. The next frequency bin was 3.125 - 6.25 kHz and there were 8192 wavelet coefficients. The third frequency bin 1.562 - 3.125 kHz contained 4096 wavelet coefficients. Towards the lower frequency bands, each successive frequency bandwidth is reduced by half.

large original files contain whale or porpoise sounds not both. Several data sets of length 32768 samples corresponding approximately to 1.3 seconds, were extracted from these large files. These data sets which contained mammal sounds mixed with background noise, were used for training and testing.

3.2   Projections on Wavelet Space

As a first step in our approach, we choose to project each of the sound vectors on an orthonormal wavelet basis. Since the sound files are sequences of discrete numbers, we adopt the compactly supported wavelets Daubechies 4 [7], which are based on discrete-time filter banks. Let $f = \{f_k\}_{k=0}^{K-1}$ be the discrete version of the input signal $f(t)$ of length $K = 2^n$. In the fast discrete wavelet transform, the signal f is first decomposed into low and high frequency bands by the convolution-decimation (subsampling by two) operations of f with the pair of a low-pass filter $G = \{g_k\}_{k=0}^{L-1}$ and a high-pass filter $H = \{h_k\}_{k=0}^{L-1}$. The filters G and H satisfy the orthogonality conditions:

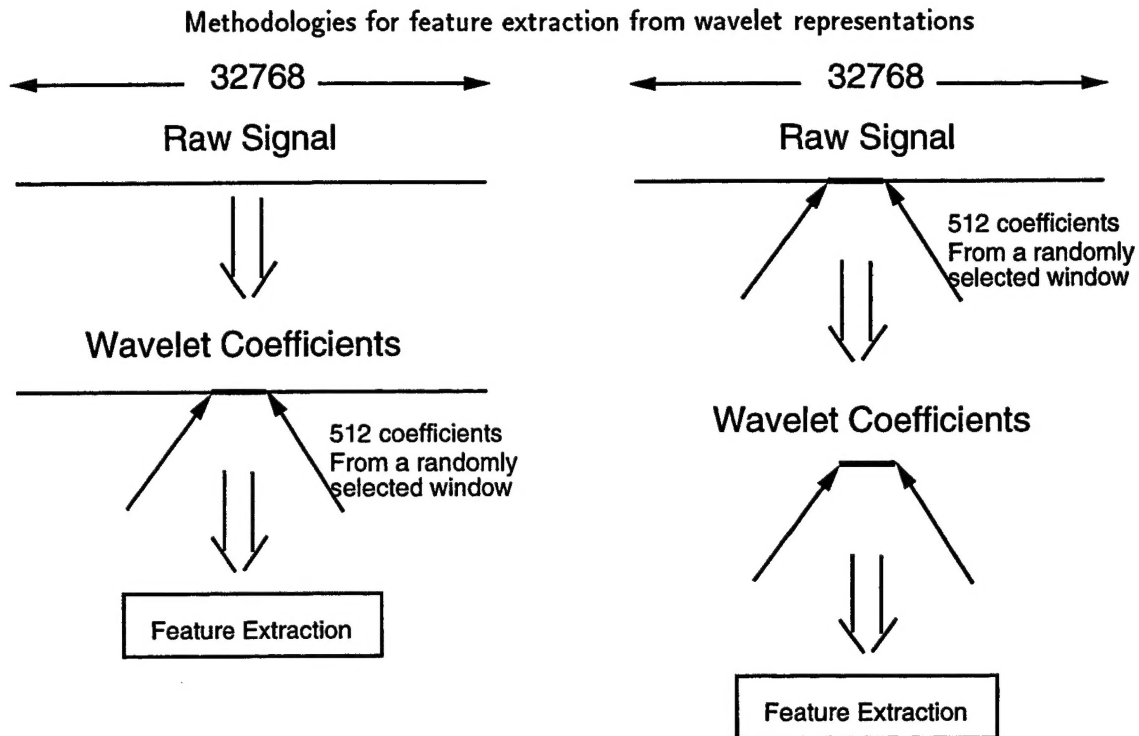$$GH^* = HG^* = 0, \quad \text{and} \quad G^*G + H^*H = I.$$

5

Methodologies for feature extraction from wavelet representations

$\longleftarrow$ —— 32768 —— $\longrightarrow$ $\longleftarrow$ —— 32768 —— $\longrightarrow$

Raw Signal **Raw Signal**

512 coefficients
From a randomly
selected window

Wavelet Coefficients

512 coefficients
From a randomly
selected window

Wavelet Coefficients

| Feature Extraction |

| Feature Extraction |

Figure 2: Application of BCM and PCA feature extraction to the wavelet representation; On the left, the raw signal with 32768 components was run through the Daubechies 4 discrete wavelet transform. From right to left are the different levels of hierarchy (16385-32768, 8193-16384, 4907-8192, etc.) correspond respectively to frequency bandwidths 6.25-12.5 KHz, 3.12-6.25 KHz, 1.56-3.12 Khz. The feature extraction method is trained on *randomly selected* 512 consecutive samples from this wavelet representation. Thus features could develp for any time/frequency combination. On the right, a segment of 512 samples from raw signal was first randomly selected, then converted to 512 wavelet coefficient and used for feature extraction.

G and H are called Quadrature Mirror Filters (QMFs). The QMFs allow *perfect reconstruction*. The decomposition process continues iteratively on the resulting low frequency bands and each time the high frequency bands are left intact. The iteration stops with one low frequency coefficient and one high frequency coefficient. As a result, the frequency axis is partitioned *smoothly* and *dyadically* finer and finer toward the low frequency region. On the time-frequency (phase), the signal is decomposed in an octave-band fashion (Figure 1). The entire phase plane is covered by disjoint cells of equal area which we call the Heisenberg cells. The uncertainty principle can be interpreted as a rectangular cell located around (t, f) that represents an uncertainty region associated with (t, f). The total number of cells is equal to the dimension of the input vector. Each cell is shaded in proportion to the amplitude of the corresponding wavelet coefficient. It is clear that this type of gray scale quantization procedure of cells conforms with the uncertainty principle.

3.3 Construction of Training Examples

We applied the wavelet transform to several porpoise and whale signals, each of which has a length of 32768 samples and a sampling rate of 25 kHz. Two different approaches to construct the training data were used. The more conventional one is described in Figure 2 (right); Here, we randomly choose small chunks of acoustic signal (512 consecutive samples) and apply wavelet analysis to get a new representation of this 512 dimensional data. Then we extract 10 features from the wavelet representation. The less conventional method is described in Figure 2 (left); Here, we first transform the full 32768 samples of the raw signal into a wavelet representation (details of the representation are in Figure 1). The two dimensional representation is then converted into a single 32768 dimensional vector. From this vector we randomly choose a chunk of 512 samples starting at a random location and use this 512-dimensional vector for feature extraction.
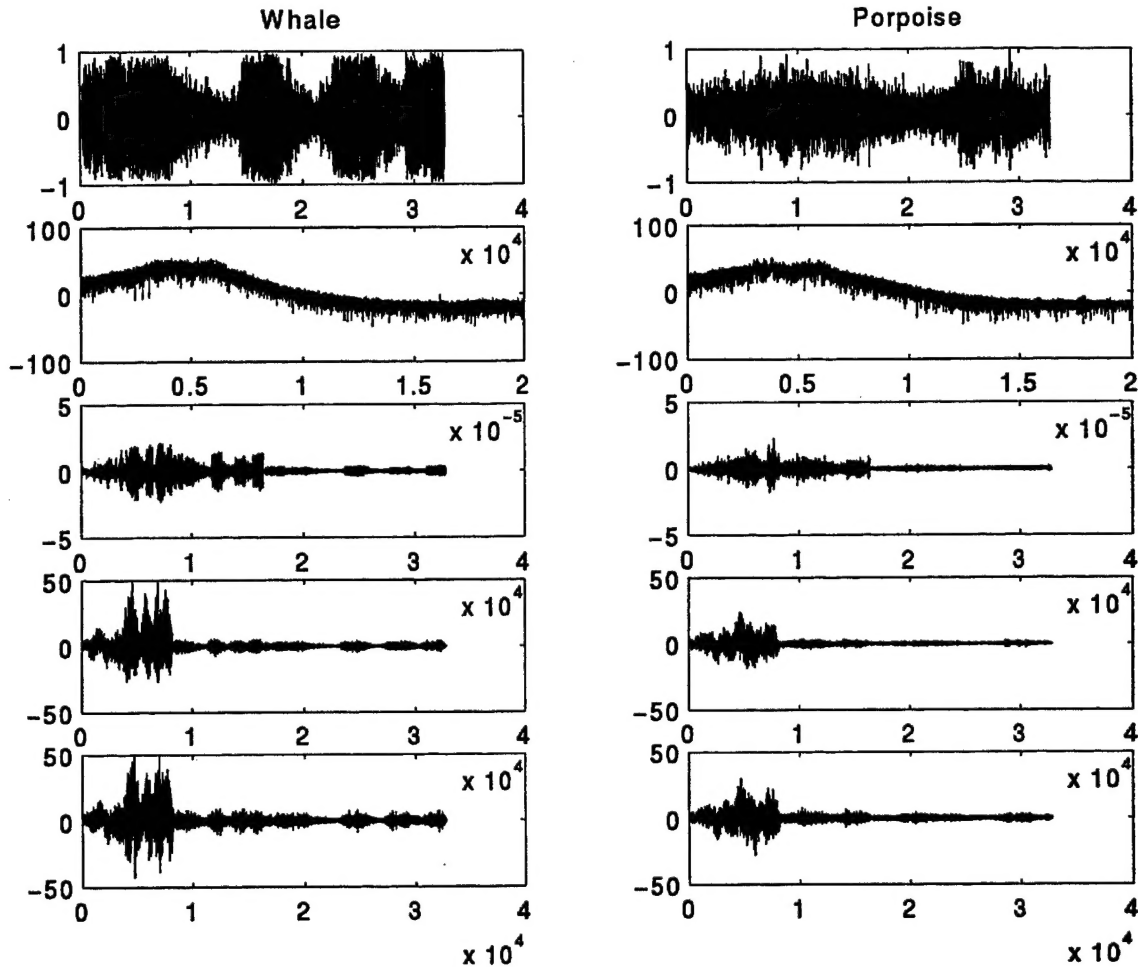
Figure 3: Various representations for the acoustic signal based on different preprocessing methods; At the top is the raw singal: 32768 consecutive samples representing approximately 1.3 sec of signal sampled at a rate of 25 kHz (horizontal axis represents time). Below, appears the Fourier representation of the signal (horizontal axis represents frequency). Note that while the raw signal does show some differences between a Porpoise signal and a Whale signal, the Fourier representations are very similar, indicating the difficulty of the classification problem. The panel below shows a wavelet representation of the signal (horizontal axis represents time and frequency). This one dimensional signal is a concatanation of time an frequency infomation (see Figure 3) so that the low frequency coefficients with low temporal resolution appear at the left, followed by high frequency with higher temporal resolution. It can be seen that the high frequency part carries less infomation compared with the lower frequency part. This fact is emphasized in the next two panels where the convolution of two BCM neurons with the wavelet signals are depicted (horizontal axis is the same as in the wavelet representation). It is clear that BCM found discriminating information in the low frequency range, at a frequency band of 1.562 − 3.125 kHz. One can then view the BCM neuron as a matched (nonlinear) filter designed to increase discrimination between the signals.

The next step of our approach was to train the k BCM neurons on the wavelet transformed data

to produce k stable weight vectors. We used here 10 BCM neurons which were connected and form a network with lateral inhibition. Each neuron was represented by one weight vector of dimension 512. The neurons were trained simultaneously on wavelet transformed signals of porpoises and whales. It took several hundred thousand iterations to converge to 10 fixed points.

Figure 3 presents various processings of the acoustic signals. There are 32768 consecutive measurements of the raw data (top panel) a Fourier representation (which looks very similar for both signals) a wavelet representation of the same signal and a convolution with two BCM neurons (bottom two panels). It can be seen that the convolution between the BCM and wavelet representation of the whale signals, indicates that the BCM neurons (all 10 of the network) respond only within the frequency bandwidth of 1.562 - 3.125 kHz at different time locations. There is no responses in the porpoise cases.

4 Classification results

We have used 300 examples of whale signals and 300 examples of porpoises for the training of the classifier. Each example was in a vector form with 10 components representing 10 features extracted by the feature extraction network. The features were computed using the two methods outlined in Section 3.3

A feed-forward neural network with 10 input nodes was used as a classifier. The architecture of the network consisted of one hidden layer with 8 nodes and one output node. The network was trained to high ninety percent correct classification.

When using the large wavelet representation for feature extraction, we have noticed that classification performance could be improved if we do not train the classifier from signals that were taken from the same frequency band (for both species). While this may sound odd, it is actually very reasonable and demonstrates a unique property of the BCM feature extraction (see Section 5); The selective response of BCM neurons to a specific frequency band was mainly seen for the Whale signals, due to the feature vectors becoming orthogonal to the class of Porpoise sounds. The orthogonality to the other class of signals caused difficulties for the classifier to converge, as there was no *error* signal. We have therefore used the frequency bin 1.562 - 3.125 kHz, which contains 4096 wavelet coefficients for the Porpoise signal. During *testing* of the classifier, only the *same* frequency band was used for both species (since one does not know apriori to what animal the signal belongs to). Thus the "Different freq. bins" referred to in Table 1 corresponds to the training methodology only.

The results presented in Tables 1 and 2, are for test data that was recorded from different oceans thus, representing a different acoustic environment and possibly different specie types. These results are therefore not comparable to results shown in [18] where training and testing was done from the same geographical location and possibly same animal. We have performed such analyses as well and got results in the range of 95%–100% correct classification.

4.1 Importance of BCM feature extraction

In this case we have studied feature extraction from the compactly supported wavelet Daubechies 4 representation. We have compared the BCM feature extraction to PCA feature extraction from this representation and tested whether the squared coefficients were more informative than the coefficients themselves, as is often assumed. It turned out that the squared coefficients which

Classification results: Wavelet analysis on 32768 dimensions

	Porpoise	Sperm whale
PCA from squared wavelet	76.7	32
BCM from same freq. bins (orig. wavelet)	92	74
BCM from different freq. bins (orig. wavelet)	96	88
BCM from same freq. bins (squared wavelet)	100	81
BCM from different freq. bins (squared wavelet)	100	91

Table 1: Percent correct classification using PCA and BCM feature extraction from Daubechies 4 basis representation. Results are presented based on features extraction directly from the coefficients or from the square of the coefficients (the energy). Results are also presented for training the classifier based on features extracted by BCM from the whole wavelet representation, namely from all frequency bands, or based on features extracted only from locations BCM was selective to (see text for details). 10 features were extracted in each of these methods,

correspond to the energy in a particular time/frequency location are more informative as is seen in Table 1. Most importantly, the BCM feature extraction outperforms PCA feature extraction from this representation. PCA (Principal Components Analysis) is much used in signal processing as it is very simple to apply and extracts second order statistics from the data which is sufficient for many applications [8]. As is seen in Table 1, the performance of PCA here is worse, suggesting that there is higher order statistics involved in the structure exploration.

4.2 Importance of the wavelet representation

The Fourier representation of the data was not useful for discrimination as it was very similar for both species (Figure 3, second panel from top). The usefulness of wavelet representation for classification of underwater sounds has been extensively studied and briefly reviewed in Section 1.1. We have thus not attempted to compare classification performance based on a wavelet to performance based on other representations. However, since we have been using a novel feature extraction method for these signals, we evaluated the performance of the BCM feature extraction based on the wavelet representation and compared it to performance on feature extraction via BCM from the raw signal.

Table 2 presents classification results from the more conventional way of extracting features from this data, a method that allows comparison with the Local Discriminant Basis search [3]. The preprocessing used is described in Figure 2 (Right). The first raw represents results of feature extraction taken directly from the raw signal, namely choosing randomly 512 consecutive measurements from the raw signal and using them as input to the BCM feature extraction. The high sensitivity to the Whale signal is in contrast to the high sensitivity of the other methods to the Porpoise signal. This suggests a possible combination between these two signal representations in the future. We have also compared two different wavelet representations: the compactly supported

Classification results: Wavelet analysis on 512 dimensions

	Porpoise	Sperm Whale
BCM applied on raw signals	32	95
LDB on wavelet packet	98	51
Highest energ. from Daub. 4	72	47
BCM extraction from Daub. 4	99	76

Table 2: Percent correct classification based on various signal representations (see text for details). BCM applied to the raw data is performed by extracting 10 features while training on randomly chosen sequential chunks of 512 samples from the 32768 sample raw data. LDB on wavelet packet extracts 10 best discriminant basis functions based on Coifman's algorithm [5]. Highest energy corresponds to extracting 5 highest energy coefficients with their location (10 features total) from Daubechies 4 basis. The last row represents classification performance on 10 BCM features extracted from Daubechies 4 basis representation.

wavelet Daubechies[2] 4 [7] and the wavelet packet representation with the "local discriminating basis" (LDB) feature extraction of Coifman and Saito [5]. LDB gets the closest results to classification from BCM features.

5   Conclusions

We have shown that feature extraction from a wavelet representation has a profound effect on the classification results. While wavelet representations are certainly more appropriate for these acoustic signals, the detailed resulting representation is not directly appropriate for classification, as it is too big. We have shown the useful properties of an efficient non-linear feature extraction method for classification from wavelet representations.

The BCM feature extraction which performs non-linear unsupervised dimensionality reduction, was found to be more practical than unsupervised principal components on one hand and supervised discriminant pursuit on the other. Rather than looking for the projections that minimize the ratio of the within-class distance vs. the between-class distance (as is done in discriminant analysis) [9], BCM looks for a direction that is mostly orthogonal to one group of signals (without knowing if they belong to the same class or not) while retaining selectivity to the other set of signals.

We have also demonstrated the ability of this method to extract features from the huge full-signal wavelet representation. This is a unique feature which can not be performed by linear discrimination [3]. Classification based on this feature extraction achieved outstanding results on test data that was recorded at the same environment as well as data that was remotely recorded.

---

[2]The third raw represents classification from the 10 highest energy coefficients of the wavelet representation.

Acknowledgments

References

[1] E. L. Beienenstock, L. N Cooper, and P. W. Munro. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal Neuroscience*, 2:32–48, 1982.

[2] R. E. Bellman. *Adaptive Control Processes*. Princeton University Press, Princeton, NJ, 1961.

[3] J. Buckheit and D. L. Donoho. Improved linear discrimination using time-frequency dictionaries. Stanford university technical report, 1995.

[4] E. E. Clothiaux, L. N Cooper, and M. F. Bear. Synaptic plasticity in visual cortex: Comparison of theory with experiment. *Journal of Neurophysiology*, 66:1785–1804, 1991.

[5] R. R. Coifman and N. Saito. Constructions of local orthonormal bases for classification and regression. *Comptes Rendus Acad. Sci. Paris, Serie I*, 319(2):191–196, 1994.

[6] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Info. Theory*, 38(2):713–719, 1992.

[7] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure and Appl. Math.*, XLI:909–996, 1988.

[8] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Wiley, New York, 1973.

[9] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

[10] J. H. Friedman. Exploratory projection pursuit. *Journal of the American Statistical Association*, 82:249–266, 1987.

[11] P. J. Huber. Projection pursuit. (with discussion). *The Annals of Statistics*, 13:435–475, 1985.

[12] Q. Huynh, W. Greene, and J. Impagliazzo. Feature extraction and classification of underwater acoustic signals. In O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt, editors, *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, pages 183–188. Kluwer, Dordrecht, The Netherlands, 1995.

[13] Nathen Interator. Feature extraction using an unsupervised neural network. *Neural Computation*, 4:98–108, 92.

[14] N. Intrator and L. N Cooper. Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks*, 5:3–17, 1992.

[15] N. Intrator and J. I. Gold. Three-dimensional object recognition of gray level images: The usefulness of distinguishing features. *Neural Computation*, 5:61–74, 1993.

[16] N. Intrator, J. I. Gold, H. H. Bülthoff, and S. Edelman. Three-dimensional object recognition using an unsupervised neural network: Understanding the distinguishing features. In Y. Feldman and A. Bruckstein, editors, *Proceedings of the 8th Israeli Conference on AICV*, pages 113–123. Elsevier, 1991.

[17] N. Intrator, D. Reisfeld, and Y. Yeshurun. Face recognition using a hybrid supervised/unsupervised neural network. *Pattern Recognition Letters*, 17:67–76, 1996.

[18] R. E. Learned and A. S. Willsky. A wavelet packet approach to transient signal classification. *Appl. Comput. Harmonic Anal.*, 2(3):265–278, 1995.

[19] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel Distributed Processing*, volume 1, pages 318–362. MIT Press, Cambridge, MA, 1986.